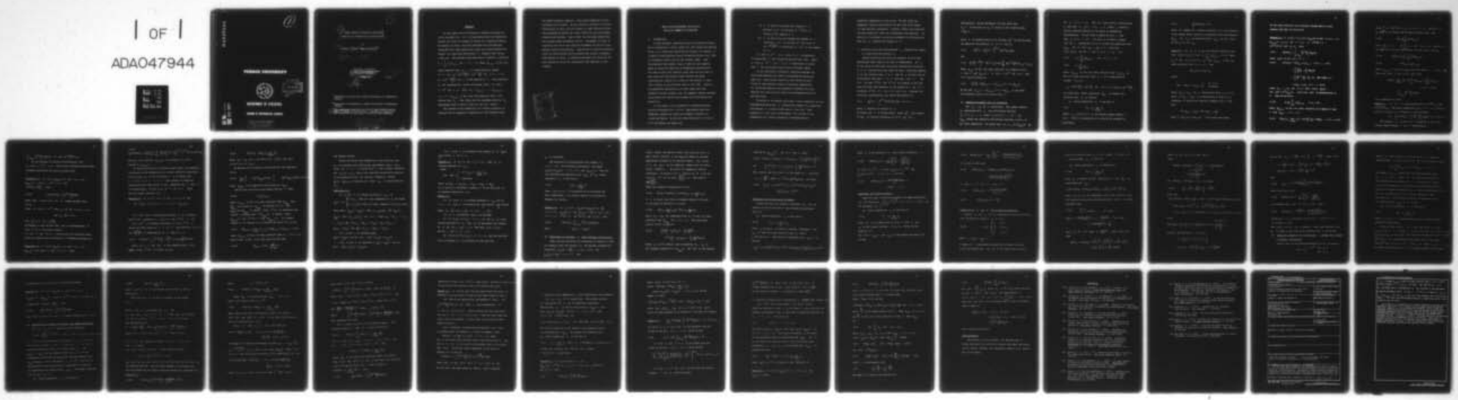
AD-A047 944 PURDUE UNIV LAFAYETTE IND DEPT OF STATISTICS F/G 12/1
SUBSET SELECTION PROCEDURES FOR RESTRICTED FAMILIES OF PROBABIL--ETC(U)
NOV 77 S S GUPTA, M LU N00014-75-C-0455
UNCLASSIFIED MIMEOGRAPH SER-518 NL

1 OF 1
ADA047944



END
DATE
FILMED
1-78
DDC

PURDUE UNIVERSITY

DEPARTMENT OF STATISTICS

DIVISION OF MATHEMATICAL SCIENCES

(1)

(6)

Subset Selection Procedures for Restricted Families of Probability Distributions,[†]

by

(10) Shanti S. Gupta[*] ~~and~~ Ming-Wei Lu[**]

(9) Technical rept.,

Department of Statistics
Division of Mathematical Sciences
(14) Mimeograph Series #518

(11) November 1977

(12) 41p.

[*]Department of Statistics, Purdue University, W. Lafayette, Indiana

[**]Department of Mathematics, Indiana University, Bloomington, Indiana

[†]This research was supported by the Office of Naval Research contract (15) N00014-75-C-0455 at Purdue University. Reproduction in whole or in part is permitted for any purpose of the United States Government.

291 730

LB

ABSTRACT

In this paper we are interested in studying multiple decision procedure for $k(k \geq 2)$ populations which are themselves unknown but which one assumed to belong to a restricted family. We propose to study a selection procedure for distributions associated with these populations which are convex-ordered with respect to a specified distribution G assuming there exists a best one. The procedure described here is based on a statistic $T_i = \sum_{j=1}^{r} a_j X_{i;j,n}$ for $i = 1,\ldots,k$ where $X_{i;j,n}$ is the j-th order statistic from F_i, r is a fixed positive integer $(1 \leq r \leq n)$, $a_j = gG^{-1}\left(\frac{j-1}{n}\right) - gG^{-1}\left(\frac{j}{n}\right)$ for $j = 1,\ldots,r-1$, $a_r = gG^{-1}\left(\frac{r-1}{n}\right)$ and g is the density of G. This statistic T_i was considered by Barlow and Doksum (1972). If $G(x) = 1 - e^{-x}$ for $x > 0$, then $nT_i = X_{i;1,n} + \ldots + X_{i;r-1,n} + (n-r+1)X_{i;r,n}$ is the total life statistic until r-th failure from F_i. This shows that the procedure based on T_i generalizes Patel's result (1976) for the IFR family.

The infimum of the probability of a correct selection is obtained and the asymptotic expression is also obtained using the subset selection approach. Some other properties of this procedure are discussed. We also study the asymptotic relative efficiencies of this rule with respect to some selection procedures proposed by Barlow and Gupta (1969) for the star-shaped ordered distributions, Gupta (1963) for the gamma populations with unknown shape parameters and etc. An example is given to illustrate the use of the selection procedure for the two independent uniform distributions. Application to quantile selection rules for distributions convex ordered with respect to Weibull distribution is given. A selection procedure for selecting the best population using the indifference zone approach is also studied.

SUBSET SELECTION PROCEDURES FOR RESTRICTED FAMILIES OF PROBABILITY DISTRIBUTIONS

1. Introduction

In many problems, especially those in reliability theory, one is interested in using a model for life length distribution which is not completely specified but belongs, for example, to a family of distributions having increasing failure rate (IFR), or increasing failure rate on the average (IFRA). Such distributions form special cases of what are now commonly known as restricted families of probability distributions. The idea of using such families stems from the fact that in many cases the experimenter cannot specify the model (distribution) exactly but is able to say whether it comes from a family of distributions such as IFR, IFRA. Families of probability distributions of these types have been studied by several authors, see, for example, Barlow, Marshall and Proschan [4], Barlow and Proschan [5,6] and Barlow and Doksum [1] .

In this paper we are interested in studying multiple decision procedures for k $(k \geq 2)$ populations which are themselves unknown but which are assumed to belong to a restricted family. We now give some definitions of interest to us (see Barlow and Gupta [3]).

(i) F is said to be convex with respect to G (written $F \underset{c}{<} G$) if and only if $G^{-1}F(x)$ is convex on the support of F.

(ii) F is said to be star-shaped with respect to G (written $F \underset{*}{<} G$) if and only if $F(0) = G(0) = 0$ and $\frac{G^{-1}F(x)}{x}$ is increasing in $x \geq 0$ on the support of F.

If $G(x) = 1 - e^{-x}$, $x \geq 0$, then $F \underset{c}{<} G$ is equivalent to saying that F has increasing failure rate (IFR). Again if $G(x) = 1 - e^{-x}$, $x \geq 0$, $F \underset{*}{<} G$ is equivalent to saying that F has increasing failure rate on average (IFRA).

In the statistical literature, selection problems for restricted families were first investigated by Barlow and Gupta [3]. Some further results in this direction and a review of some important results concerning inequalities for restricted families and problems of inference for such families have been given by Gupta and Panchapakesan [10,11] and Patel [15].

In Section 2, we propose and study a subset selection rule for distributions which are $\underset{c}{<}$ ordered with respect to a specified distribution G assuming there exists a best one. Some properties of this rule are discussed. The infimum of the probability of a correct selection is obtained and an asymptotic expression is also given. We also study the asymptotic relative efficiencies of this rule with respect to some selection procedures. Section 3 deals with selecting the best population using the indifference zone approach. In Section 4, we propose a selection procedure for distributions that are $<_*$ ordered with respect to G .

2. Selection rules for distributions $<_c$ ordered with respect to a specified distribution G .

Before discussing the selection problem, we give some preliminary known results for sake of completeness. Let $\mathcal{G}$ be the class of absolutely continuous distribution functions G on R with positive and right-(or left-) continuous density g on the interval where $0 < G < 1$ and let $\mathcal{F}$ be the class of distribution functions F in $\mathcal{G}$ such that $F(0) = 0$. For $G \in \mathcal{G}$, we take $G^{-1}(0)$ and $G^{-1}(1)$ to be equal to the left hand and right hand endpoints of the support of G. For $F \in \mathcal{F}$, we define $F^{-1}(0) = 0$. For $F \in \mathcal{F}$ and $G \in \mathcal{G}$, consider the following transformation (see Barlow and Doksum [1])

$$(2.1) \qquad H_F^{-1}(t) = \int_0^{F^{-1}(t)} g[G^{-1}F(u)]du \, , \; 0 \le t \le 1 \, ,$$

where g denotes the density of G .

We assume that G is always fixed. Since H_F^{-1} (the inverse of H_F) is strictly increasing on $[0,1]$, H_F is a distribution. Barlow and Doksum [1] have shown that $F \underset{c}{<} G$ if and only if H_F is convex on the interval where $0 < H_F < 1$.

Since G is assumed known we can estimate H_F^{-1} by substituting the empirical distribution F_n of F; that is

(2.2) $$H_n^{-1}(t) = H_{F_n}^{-1}(t) = \int_0^{F_n^{-1}(t)} g[G^{-1}F_n(u)]du$$

and

(2.3) $$H_n^{-1}(\tfrac{r}{n}) = \int_0^{X_{r,n}} g[G^{-1}F_n(u)]du = \sum_{i=1}^{r} g[G^{-1}(\tfrac{i-1}{n})](X_{i,n} - X_{i-1,n})$$

where $X_{i,n}$ is the i-th order statistic in a sample of size n from F and $X_{0,n} \equiv 0$. If $G(x) = 1 - e^{-x}$ for $x \geq 0$, then (2.3) can be written as

(2.4) $$H_n^{-1}(\tfrac{r}{n}) = \tfrac{1}{n}[X_{1,n} + \ldots + X_{r-1,n} + (n-r+1)X_{r,n}] .$$

We say that $X_{1,n} + \ldots + X_{r-1,n} + (n-r+1)X_{r,n}$ is the total life statistic until r-th failure from F.

(A) <u>Selection procedure and its properties</u>

Let $\pi_1, \ldots, \pi_k$ be k populations. The random variable X_i associated with π_i has distribution function F_i, $i = 1, 2, \ldots, k$, where $F_i \in \mathfrak{F}$ $(i = 1, \ldots, k)$. Let $F_{[k]}$ denote the cumulative distribution function (c.d.f.) of the "best" population. We assume that (a) $F_i(x) \geq F_{[k]}(x)$ for all x, $i=1, \ldots, k$ and (b) there exists a distribution G such that $F_i \underset{\sim}{<} G$, $i=1, \ldots, k$, where $\underset{\sim}{<}$ denotes a partial ordering relation on the space of probability distributions. We are given a sample of size n from each π_i $(i=1, \ldots, k)$. Our goal is to select a subset from the k populations so as to include the population with $F_{[k]}$. Let $\Omega = \{\underline{F} = (F_1, \ldots, F_k) : \exists$ a j such that $F_i(x) \geq F_j(x)$ for all x and $i=1, 2, \ldots, k\}$. Let

(2.5) $$T_i = \sum_{j=1}^{r} a_j X_{i;j,n} \quad \text{for } i=1, \ldots, k .$$

(2.6) $$T = \sum_{j=1}^{r} a_j Y_{j,n}$$

where $X_{i;j,n}$ is the j-th order statistic from $F_i, Y_{j,n}$ is the j-th order statistic from G, r is a fixed positive integer $(1 \leq r \leq n)$,

$$a_j = gG^{-1}\left(\frac{j-1}{n}\right) - gG^{-1}\left(\frac{j}{n}\right) \quad \text{for } j=1,\ldots,r-1 \text{ and } a_r = gG^{-1}\left(\frac{r-1}{n}\right) .$$

For selecting a subset containing $F_{[k]}$, we propose the selection rule R_1 as follows:

R_1: Select population π_i if and only if

(2.7) $$T_i \geq c_1 \max_{1 \leq j \leq k} T_j$$

where $c_1 = c_1(k, P^*, n, r)$ is the largest number between 0 and 1 which is determined as to satisfy the probability requirement

$$\inf_{\Omega} P\{CS|R_1\} \geq P^* \tag{2.8}$$

where CS stands for a correct selection, i.e., the selection of any subset which contains the population with distribution $F_{[k]}$. Let $T_{(i)}$ be associated with $F_{[i]}$ and let $W_i(x)$ be the c.d.f. of $T_{(i)}$.

Lemma 2.1. Let F_1, F_2 be two distribution functions such that $F_1(x) \geq F_2(x)\ \forall x$ and $T_i = \sum_{j\in\Delta} b_j X_{i;j,n}$ $i=1,2$, where $b_j > 0$ for $j \in \Delta$, $\Delta \subset \{1,2,\ldots,n\}$ and $X_{i;j,n}$ is the j-th order statistic from F_i, $i=1,2$, then

$$P[T_1 \leq x] \geq P[T_2 \leq x] .$$

Proof.

$$\text{Let } \psi(X_{i1},\ldots,X_{in}) = \begin{cases} 1 & \text{if } T_i \geq x \\ 0 & \text{otherwise} \end{cases}$$

where $X_{i1},\ldots,X_{in}$ are n observations from F_i $(i=1,2)$. Since $\psi(X_{i1},\ldots,X_{in})$ is nondecreasing in each of its arguments, it follows by induction (Lehmann [11] P. 112) that

$$E\psi(X_{11},\ldots,X_{1n}) \leq E\psi(X_{21},\ldots,X_{2n})$$

That is $P[T_1 \geq x] \leq P[T_2 \geq x]$. This proves the lemma.

We now state and prove the following theorem which is more general than that of Patel [15].

Theorem 2.1. If $F_i \in \mathfrak{F}$, $G \in \mathcal{G}$, $F_i(x) \geq F_{[k]}(x)$ $\forall x$ and $i = 1,2,\ldots,k$, $F_{[k]} \underset{c}{<} G$, $a_j \geq 0$ for $j = 1,2,\ldots,r$, $G^{-1}(0) \leq 0$, $g\, G^{-1}(0) \leq 1$ and $a_r \geq c_1$, then

$$(2.9) \qquad P[CS|R_1] \geq \int_{G^{-1}(0)}^{\infty} G_T^{k-1}\left(\frac{x}{c_1}\right) dG_T(x)$$

where $G_T(x)$ is the c.d.f. of T.

Proof. $\quad P[CS|R_1] = P[T_{(k)} \geq c_1 T_{(i)}\,,\ i = 1, \ldots, k-1]$

$$= \int_0^{\infty} \prod_{i=1}^{k-1} W_i\left(\frac{x}{c_1}\right) dW_k(x)$$

$$\geq \int_0^{\infty} W_k^{k-1}\left(\frac{x}{c_1}\right) dW_k(x) \quad \text{(By Lemma 2.1)}$$

$$= P[Z_k \geq c_1 Z_j,\ j = 1, \ldots, k-1]$$

where $Z_1, \ldots, Z_k$ are i.i.d. with c.d.f. $W_k(x)$.

Let $\varphi(x) = G^{-1}F_{[k]}(x)$. Note that $\varphi(x)$ is nondecreasing in x. Also we can write

$$(2.10) \qquad Z_i \overset{=}{\underset{st}{=}} \sum_{j=1}^{r} a_j X^*_{i;j,n} \qquad i = 1,\ldots,k\,,$$

where $X^*_{i;j,n}$ is the j-th order statistic in a sample of size n from $F_{[k]}$, $i = 1,\ldots,k$.

$$(2.11) \qquad P[Z_k \geq c_1 \max_{1\leq j\leq k} Z_j] = P[\varphi(\frac{1}{c_1} Z_k) \geq \varphi(Z_i)\,,\ i = 1,\ldots,k-1]\,.$$

Since $\sum_{j=1}^{r} a_j = gG^{-1}(0) \le 1$, $a_j \ge 0 \ \forall j = 1,\dots,r$, and $\varphi(0) \le 0$, by Lemma 4.1 of Barlow and Proschan [5] and (2.10), then

$$\varphi(Z_i) \underset{st}{\le} \sum_{j=1}^{r} a_j \varphi(X^*_{i;j,n}) . \tag{2.12}$$

Since $\frac{1}{c_1} a_r \ge 1$, $\frac{1}{c_1} \sum_{j=1}^{r} a_j \ge 1$ for $i = 1,\dots,r$, and $\varphi(0) \le 0$ by Lemma 4.3 of Barlow and Proschan [5] and (2.10), we have

$$\varphi(\frac{1}{c_1} Z_k) \underset{st}{\ge} \frac{1}{c_1} \sum_{j=1}^{r} a_j \varphi(X^*_{k;j,n}) . \tag{2.13}$$

$$\varphi(X^*_{i;j,n}) \underset{st}{=} Y_{i;j,n} \tag{2.14}$$

where $Y_{i;j,n}$ is the j-th order statistic from G, $i = 1,2,\dots,k$.
Thus from (2.11), (2.12), (2.13), and (2.14),

$$P[Z_k \ge c_1 \max_{1 \le i \le k} Z_i] \ge P[\sum_{j=1}^{r} a_j Y_{k;j,n} \ge c_1 \sum_{j=1}^{r} a_j Y_{i;j,n} ,$$

$$i = 1,2,\dots,k-1]$$

$$= \int_{G^{-1}(0)}^{\infty} G_T^{k-1}(\frac{x}{c_1}) dG_T(x)$$

This completes the proof.

Remark 2.1. (i) If g is nonincreasing, $g(0) \le gG^{-1}(0) \le 1$ and $gG^{-1}\left(\frac{r-1}{n}\right) \ge c_1$, then these conditions $a_j \ge 0$, $j = 1,\dots,r$, $G^{-1}(0) \le 0$, $gG^{-1}(0) \le 1$ and $a_r \ge c_1$ in Theorem 2.1 are satisfied.

(ii) $\inf_{\Omega} P[CS|R_1] = \int_0^{\infty} G_T^{k-1}\left(\frac{x}{c_1}\right) dG_T(X)$ if $G^{-1}(0) = 0$.

The constant $c_1 = c_1(k,P^*,n,r)$ satisfying (2.8) is the largest number between 0 and 1 determined by

$$\int_{G^{-1}(0)}^{\infty} G_T^{k-1}\left(\frac{x}{c_1}\right) dG_T(x) \geq P^* \quad \text{and} \quad gG^{-1}\left(\frac{r-1}{n}\right) \geq c_1 \; .$$

We now consider two specific distributions $G(x)$.

If $G(x) = 1 - e^{-x}$, $x \geq 0$, then we have following result which slightly generalizes the result of Patel [15].

Corollary 2.1. If $F_i(x) \geq F_{[k]}(x)$ $\forall x$ and $i = 1,\ldots,k$, $F_{[k]} \underset{c}{<} G$, $G(x) = 1 - e^{-x}$, $x > 0$ and $n \geq \max\{r, \frac{r-1}{1-c_1}\}$, then

$$\inf_{\Omega} P[CS|R_1] = \int_0^{\infty} H^{k-1}\left(\frac{x}{c_1}\right) dH(x) \tag{2.15}$$

where $H(x)$ is the c.d.f. of a χ^2 random variable with $2r$ d.f.

Proof. If $G(x) = 1 - e^{-x}$ then $a_j = \frac{1}{n}$ for $j = 1,2,\ldots,r-1$

and $a_r = \frac{1}{n}(n-r+1)$.

Also $\frac{1}{c_1} a_r \geq 1$ iff $n \geq \frac{r-1}{1-c_1}$.

By Theorem 2.1 and the fact that $2nT$ is distributed as χ^2 with $2r$ d.f., the result follows.

If $G(x) = x$ for $0 < x < 1$, then we have the following result which is a special case of Theorem 2.1 of Barlow and Gupta [3].

Corollary 2.2. If $F_i(x) \geq F_{[k]}(x)$ $\forall x$ and $i = 1,\ldots,k$, $F_{[k]} \underset{c}{<} G$ and $G(x) = x$ for $0 < x < 1$, then

(2.16)

$$\inf_{\Omega} P[CS|R_1] = n\binom{n-1}{r-1}\int_0^{\infty} \left[\sum_{i=r}^{n} \binom{n}{i}\left(\frac{x}{c_1}\right)^i\left(1-\frac{x}{c_1}\right)^{n-i}\right]^{k-1} x^{r-1}(1-x)^{n-r}\,dx \;.$$

Actually, the condition $F_{[k]} \underset{c}{<} G$ in Corollary 2.2 can be relaxed to $F_{[k]} \underset{*}{<} G$.

We state and prove the following theorem about the asymptotic evaluations of the probability of a correct selection associated with the rule R_1 in the case where r is so chosen that $r \le (n+1)\alpha < r+1$, $0 < \alpha < 1$. This amounts to selecting populations with large values of the α-quantile for α (and r) as defined above. In this case, $\frac{r}{n} \to \alpha$ as $n \to \infty$. Note that the result holds for all α .

<u>Theorem 2.2</u>. If $F_i \in \mathfrak{F}$, $G \in \mathcal{G}$ for all $i = 1,\dots,k$ and

(i) $F_i(x) \ge F_k(x)\ \forall x$, $i = 1,\dots,k$, $F_{[k]} \underset{c}{<} G$,

(ii) $G(x)$ has a differentiable density g in a neighborhood of its α-quantile η_α , $g(\eta_\alpha) \ne 0$ and $G^{-1}(0) \le 0$, and

(iii) gG^{-1} is uniformly continuous on $[0,1]$, $G^{-1}(x)$ is convex and there exists an ξ , $0 < \xi < 1$, such that for $\xi \le y < 1$, and $\frac{gG^{-1}(y)}{1-y}$ is nondecreasing in y , then as $n \to \infty$

$$(2.17) \qquad P[CS|R_1] \ge \int_{-\infty}^{\infty} \Phi^{k-1}\left[\frac{x}{c_1} + \frac{1-c_1}{c_1}\,\eta_\alpha\, g(\eta_\alpha)\left(\frac{n}{\alpha\bar{\alpha}}\right)^{1/2}\right] d\Phi(x)$$

where $\bar{\alpha} = 1 - \alpha$ and $\Phi(x)$ is the standard normal c.d.f.

<u>Proof</u>. Since $G^{-1}(x)$ is convex, we have

$$P[CS|R_1] \geq P[Z_k \geq c_1 \max_{1\leq j\leq k} Z_j] \tag{2.18}$$

where $Z_1,\ldots,Z_k$ are i.i.d. with c.d.f. $W_k(x)$ and $W_k(x)$ is the c.d.f. of $T_{(k)}$.

By Theorem (2.2) of Barlow and Van Zwet [7] and condition (iii),

$$\sup_{x\geq 0}\left| \int_0^x g[G^{-1}F_{[k]n}(u)]du - \int_0^x g[G^{-1}F_{[k]}(u)]du\right| \to 0 \text{ a.s} \tag{2.19}$$

where $F_{[k]n}$ is the empirical distribution of $F_{[k]}$.

Then we have (see Barlow and Doksum [1]), for n large,

$$Z_i \underset{st}{\cong} Y_{i;r,n} \tag{2.20}$$

where $Y_{i;r,n}$ is the r-th order statistic from $H_{F_{[k]}}$ and $H_{F_{[k]}}^{-1}$ (the inverse of $H_{F_{[k]}}$) is defined in (2.1). Now $F_{[k]} \underset{c}{<} G \Leftrightarrow H_{F_{[k]}}$ is convex. Since $G^{-1}(x)$ is increasing and convex, it follows that $G^{-1}H_{F_{[k]}}(x)$ is convex. Since $H_{F_{[k]}} \underset{c}{<} G$ and $G^{-1}(0) \leq 0$, then $H_{F_{[k]}} \underset{*}{<} G$. In a manner similar to the theorem (2.1) of Barlow and Gupta [3], we have

$$P[Y_{k;r,n} \geq c_1 \max_{1<i\leq k} Y_{i;r,n}] \geq P[Y^*_{k;r,n} \geq c_1 Y^*_{i;r,n}, \ i \neq k] \tag{2.21}$$

where $Y^*_{i;r,n}$ is the r-th order statistic from G, $i=1,\ldots,k$.

From (2.18), (2.20), (2.21) and using the fact that

$$Y^*_{i;r,n} \sim N(\eta_\alpha, \frac{\alpha\bar{\alpha}}{ng^2(\eta_\alpha)}), \tag{2.22}$$

the theorem follows.

Before we discuss some properties of the selection rule R_1, we introduce some definitions (See Santner [16]). For a given $\alpha (0 < \alpha < 1)$, we assume each F_i has a unique α-quantile. Let $F_{[i]}(x) = F_{[i]}$ denote the cumulative distribution function of the population with i-th smallest α-quantile. Define $P_{\underline{F}}(i) = P_{\underline{F}}[\pi_{(i)}$ is selected $|R]$ where $\pi_{(i)}$ is associated with $F_{[i]}$.

<u>Definition 2.1</u>.

(i) A rule R is strongly monotone in $\pi_{(i)}$ if

$$P_{\underline{F}}(i) \text{ is } \begin{cases} \uparrow \text{in } F_{[i]} \text{ when all other components of } \underline{F} \text{ are fixed} \\ \downarrow \text{in } F_{[j]} (j \neq i) \text{ when all other components of } \underline{F} \text{ are fixed.} \end{cases}$$

That means, $P_{\underline{F}_1}(i) \geq P_{\underline{F}_1^*}(i)$ when $F_{[i]} \underset{st}{\geq} F^*_{[i]}$ and $P_{\underline{F}_2}(j) \leq P_{\underline{F}_2^*}(j)$ when $F_{[j]} \underset{st}{\geq} F^*_{[j]}$ for $j \neq i$, where $\underline{F}_1 = (F_{[1]}, \ldots, F_{[i]}, \ldots, F_{[k]})$, $\underline{F}_1^* = (F_{[1]}, \ldots, F^*_{[i]}, \ldots, F_{[k]})$, $\underline{F}_2 = (F_{[1]}, \ldots, F_{[j]}, \ldots, F_{[k]})$ and $\underline{F}_2^* = (F_{[1]}, \ldots, F^*_{[j]}, \ldots, F_{[k]})$.

(ii) A rule R is monotone means $P_{\underline{F}}(i) \leq P_{\underline{F}}(j)$ for all $\underline{F} \in \Omega$ with $F_{[i]}(x) \geq F_{[j]}(x)$.

(iii) A rule R is unbiased if $P_{\underline{F}}(i) \leq P_{\underline{F}}(k)$ for all $\underline{F} \in \Omega$ with $F_{[i]}(x) \geq F_{[k]}(x)$.

(iv) A rule R is consistent with respect to Ω' means $\inf_{\Omega'} P[CS|R] \to 1$ as $n \to \infty$.

<u>Theorem 2.3</u>. If $a_i \geq 0$ for $i = 1,\ldots,r$, then R_1 is strongly monotone in $\pi_{(i)}$.

Proof.

$$\text{Let } \psi(\underline{X}) = \begin{cases} 1 & \text{if } T_{(i)} \geq c \max_{1\leq j\leq k} T_{(j)} \\ 0 & \text{otherwise} \end{cases}$$

where $\underline{X} = (X_{11},\ldots,X_{1n},X_{21},\ldots,X_{2n},\ldots,X_{k1},\ldots,X_{kn})$.

It is similar to the proof in Lemma 2.1, we can show that R_1 is strongly monotone in $\pi(i)$.

<u>Remark 2.2</u>.

(1) If a rule R is strongly monotone in $\pi_{(i)}$ for all $i = 1,\ldots,k$, then R is monotone and $\inf_{\Omega} P[CS|R] = \inf_{\Omega_0} P[CS|R]$ where $\Omega_0 = \{\underline{F} = (F_1,\ldots,F_k) \in \Omega : F_1 = \ldots = F_k\}$.

(2) If R is monotone, then it is unbiased.

(3) If $F_i(x) = F(x,\theta_i)$, $i = 1,\ldots,k$ and T_i is a consistent estimator of θ_i, then R_1 is consistent with respect to $\Omega = \{\underline{F} = (F_1,\ldots,F_k) : \exists \text{ a } j$ such that $F_i(x) \geq F_j(x)$ for all x and $i = 1,\ldots,k\}$.

(4) If $F_i \in \mathfrak{F}$, $G \in \mathcal{G}$, $F_i \underset{c}{<} G$, $i = 1,\ldots,k$ and the condition (iii) of Theorem 2.2 is satisfied, we can show that

R_1 is consistent.

The selection of the population with largest F_i $(i=1,\ldots,k)$ can be handled analogously. We assume $F_{[i]}(x) \le F_{[1]}(x)$, $i=1,\ldots,k$, and $F_{[1]} \underset{c}{<} G$. The rule for selecting the population with $F_{[1]}$ is R_2: Select population π_i if and only if

$$c_2 T_i \le \min_{1 \le j \le k} T_j \tag{2.23}$$

where $c_2 (0 < c_2 \le 1)$ is determined so as to satisfy the basic requirement. In a manner similar to the proof of Theorem 2.1, we have

<u>Theorem 2.4</u>. If $F_i, G \in \mathcal{F}$, $F_{[i]}(x) \le F_{[1]}(x)$ $\forall x$ and $i=1,\ldots,k$, $F_{[1]}(0)=0$ and $F_{[1]} \underset{c}{<} G$ and if $a_j \ge 0$ for $j=1,\ldots,r$, $G^{-1}(0) \le 0$, $gG^{-1}(0) \le 1$ and $a_r \ge c_2$, then

$$P[CS|R_2] \ge \int_{G^{-1}(0)}^{\infty} \bar{G}_T^{k-1}(c_2 x)\, dG_T(x) \tag{2.24}$$

where $\bar{G}_T(x) = 1 - G_T(x)$.

(B) <u>Efficiency of procedure</u> R_1 <u>under slippage configuration</u>.

Under the same notations and conditions of Theorem 2.2 and the comments above the Theorem 2.2, we consider slippage configuration $F_{[i]}(x) = F\left(\frac{x}{\delta}\right)$, $i = 1,2,\ldots,k-1$, and $F_{[k]}(x) = F(x)$, $0 < \delta < 1$. Let

$E(S|R)$ denote the expected subset size using the rule R. Then $E(S|R) - P[CS|R]$ is the expected number of non-best populations included in the selected subset. For a given $\varepsilon > 0$, let $n_R(\varepsilon)$ be the asymptotic sample size for which $E(S|R) - P\{CS|R\} = \varepsilon$. We define the asymptotic relative efficiency $A\,R\,E(R,R^*,\delta)$ of R relative to R^* to be the limit as $\varepsilon \to 0$ of the ratio $\frac{n_R(\varepsilon)}{n_{R^*}(\varepsilon)}$, i.e. $ARE\,(R,R^*;\delta) = \lim_{\varepsilon\to 0} \frac{n_R(\varepsilon)}{n_{R^*}(\varepsilon)}$.

Under the slippage configuration we have,

$$(2.25) \qquad E(S|R_1) = P[CS|R_1] + (k-1)P[T_{(1)} \geq c_1 \max_{i\neq 1} T_{(i)}]$$

If n is large, then from an argument similar to the one in the proof of Theorem 2.2, we have

$$(2.26) \qquad P[T_{(1)} \geq \max_{i\neq 1} T_{(i)}] \cong P[Y_1 \geq c_1 \max_{i\neq 1} Y_i]$$

where $Y_1,\ldots,Y_k$ are independent and Y_i is the r-th order statistic from $H_{F_{[i]}}$ for $i=1,\ldots,k$. The right-hand side of (2.26) is equal to

$$(2.27) \qquad \int_{-\infty}^{\infty} \Phi\left(\frac{\delta x}{c_1} - a_\alpha h(a_\alpha)\left(1-\frac{\delta}{c_1}\right)\left(\frac{n}{\alpha\alpha}\right)^{\frac{1}{2}}\right) .$$
$$\Phi^{k-2}\left(\frac{x}{c_1} - a_\alpha h(a_\alpha)\left(1-\frac{1}{c_1}\right)\left(\frac{n}{\alpha\alpha}\right)^{\frac{1}{2}}\right) d\Phi(x)$$

where c_1 is the constant used in defining R_1, a_α is the (unique) α-quantile of $H_{F_{[k]}}(x)$ and $h(x)$ is the density function of $H_{F_{[k]}}(x)$. For $k=2$ and n large,

$$(2.28)\quad E(S|R_1) - P[CS|R_1] \simeq \Phi(-h(a_\alpha)a_\alpha(1-\frac{\delta}{c_1})(\frac{n}{\alpha\bar{\alpha}})^{\frac{1}{2}}(1+\frac{\delta^2}{c_1^2})^{-\frac{1}{2}}) .$$

Let $\int_{-\infty}^{\infty} \Phi^{k-1}(\frac{x}{c_1} + (1-c_1)\eta_\alpha g(\eta_\alpha) \frac{1}{c_1} (\frac{n}{\alpha\bar{\alpha}})^{\frac{1}{2}}) d\Phi(x) = P^*$.

Now, setting the right aide of (2.28) equal to ϵ and using

$c_1 \approx 1 - \frac{2^{\frac{1}{2}}D}{n^{\frac{1}{2}}}$, where $D = \Phi^{-1}(P^*)(\alpha\bar{\alpha})^{\frac{1}{2}}/\eta_\alpha g(\eta_\alpha)$, we obtain

$$(2.29)\quad n_{R_1}(\epsilon) \approx [-(\alpha\bar{\alpha})^{\frac{1}{2}}\Phi^{-1}(\epsilon)(1+\delta^2)^{\frac{1}{2}} + \sqrt{2}\, D\delta a_\alpha h(a_\alpha)]^2 \cdot [a_\alpha^2 h^2(a_\alpha)(1-\delta)^2]^{-1} .$$

<u>Comparison with Barlow-Gupta Procedure</u>

Barlow and Gupta [3] propose a procedure R_3 , for the quantile selection problem of star-ordered distributions which is,

R_3: Select population π_i if and only if

$$(2.30)\quad T_{r,i} \geq c_3 \max_{1 \leq j \leq} T_{r,j}$$

where c_3 $(0 < c_3 < 1)$ is chosen to satisfy $P[CS|R_3] \geq P^*$ and $T_{r,i}$ is the r-th order statistic from F_i where $r \leq (n+1)\alpha < r+1$. They derive an expression for $n_{R_3}(\epsilon)$ as follows:

$$n_{R_3}(\epsilon) \approx [-(\alpha\bar{\alpha})^{\frac{1}{2}}\Phi^{-1}(\epsilon)(1+\delta^2)^{\frac{1}{2}} + \sqrt{2}\, D\delta\xi_\alpha f(\xi_\alpha)]^2[\xi_\alpha^2 f^2(\xi_\alpha)(1-\delta)^2]^{-1}$$

where f is the density of F with unique α-quantile, ξ_α.

$$\text{(2.31)} \qquad ARE(R_1, R_3; \delta) = \lim_{\epsilon \to 0} \frac{n_{R_1}(\epsilon)}{n_{R_3}(\epsilon)} = \frac{\xi_\alpha^2 f^2(\xi_\alpha)}{a_\alpha^2 h^2(a_\alpha)} .$$

If $G(x) = 1 - e^{-x}$, $x > 0$ and $F_{[1]}(x) = 1 - e^{-x/\delta}$ and $F_{[2]}(x) = 1 - e^{-x}$, $x \geq 0$, $0 < \delta < 1$, we have,

$$\text{(2.32)} \qquad ARE(R_1, R_3; \delta) = \frac{(1-\alpha)^2 \log^2(1-\alpha)}{\alpha^2} < 1$$

$$= 0.4803 , \quad \alpha = \tfrac{1}{2} .$$

Comparison with Gupta Procedure

Gupta [8] gave a selection procedure for gamma populations π_i's with densities $\frac{1}{\Gamma(a)\theta_i^r} x^{a-1} e^{-\frac{x}{\theta_i}}$ $x > 0$, $\theta_i > 0$, $i = 1, 2, \ldots, k$. The procedure R_4 is

R_4: Select population π_i if and only if

$$\text{(2.33)} \qquad \bar{X}_i \geq c_4 \max_{1 \leq j \leq k} \bar{X}_j$$

where $\bar{X}_i$ is the sample mean of size n from π_i and c_4 is the largest constant $(0 < c_4 \leq 1)$ chosen so that $P[CS \mid R_4] \geq P^*$.

For $k = 2$, $\theta_{[1]} = \delta$ and $\theta_{[2]} = 1$ (see Barlow and Gupta [3]), we have

$$(2.34)\qquad ARE(R_3,R_4;\delta) = \lim_{\varepsilon\to 0} \frac{n_{R_3}(\varepsilon)}{n_{R_4}(\varepsilon)} = \frac{a(\log\,\delta)^2 \alpha\bar{\alpha}(1+\delta^2)}{2(1-\delta)^2[\xi_\alpha f(\xi_\alpha)]^2}\ .$$

It is easy to show that

$$(2.35)\qquad ARE(R_1,R_4;\delta) = ARE(R_1,R_3;\delta)ARE(R_3,R_4;\delta)$$

$$= \left\{\frac{\sqrt{a}\,\log\,\delta\,\sqrt{\alpha\bar{\alpha}}\,\sqrt{1+\delta^2}}{\sqrt{2}\,(1-\delta)a_\alpha\;h(a_\alpha)}\right\}^2\ .$$

If $G(x) = 1 - e^{-x}$ for $x > 0$ and $a = 1$,

$$(2.36)\qquad ARE(R_1,R_4;\delta) = \frac{(1-\alpha)(1+\delta^2)\log^2\delta}{2(1-\delta)^2\alpha}\ .$$

$$(2.37)\qquad ARE(R_1,R_4;\delta\uparrow 1) = \frac{1-\alpha}{\alpha}\ .$$

<u>Comparison of</u> R_1 <u>and</u> R_5 <u>from uniform distribution</u>

Suppose π_1 and π_2 are two independent uniform populations with distribution functions F_i $(i=1,2)$.

$$(2.38)\qquad F_i(x) = \begin{cases} 0 & x < 0 \\ \dfrac{x}{\theta_i} & 0 \le x < \theta_i \\ 1 & x > \theta_i \end{cases}$$

where $\delta = \theta_{[1]} < \theta_{[2]} = 1$.

A sample of n independent observations is drawn from each of the two populations. Let T_i^* be the total life statistic until r-th failure from π_i $(i = 1,2)$ where $r \leq (n+1)\alpha < r+1$.

The procedure R_5 is given by

R_5: Select population π_i if and only if

$$(2.39) \qquad T_i^* \geq c_5 \max_{1 \leq j \leq k} T_j^*$$

where c_5 is chosen so that $P[CS|R_5] \geq P^*$. Let $T_{(i)}^*$ be associated with $\theta_{[i]}$.

$$(2.40) \quad E(S|R_5) - P[CS|R_5] = P[T_{(1)}^* \geq c_5 T_{(2)}^*] = P[T_i' \geq \frac{c_5}{\delta} T_2']$$

where T_1', T_2' are two independent total life statistic until r-th failure from uniform distribution over $(0,1)$. By Gupta and Sobel [12],

$$(2.41) \qquad \frac{T_i' - u}{\sigma} \to N(0,1) \qquad \text{as} \quad n \to \infty ,$$

where $u = \dfrac{n\alpha(2n - \alpha n + 1)}{2n+1} \approx u' = \dfrac{n\alpha(2-\alpha)}{2}$, $\sigma^2 = An$ and

$$A = \frac{\alpha(1-\alpha)(2-\alpha)^2}{4} + \frac{\alpha^3}{12} .$$

Hence $\dfrac{u}{\sigma} \approx \dfrac{u'}{\sigma} = B\sqrt{n}$ where $B = \dfrac{\alpha(2-\alpha)}{2\sqrt{A}}$. From (2.40), we have

$$\begin{aligned} E(S|R_5 - P[CS|R_5] &= P[\frac{T_1' - u'}{\sigma} \geq \frac{c_5}{\delta}(\frac{T_2' - u'}{\sigma}) + (\frac{c_5}{\delta} - 1)\frac{u'}{\sigma}] \\ &\approx P[Z_1 \geq \frac{c_5}{\delta} Z_2 + (\frac{c_5}{\delta} - 1)B\sqrt{n}] \end{aligned}$$

where Z_1, Z_2 are i.i.d. with $N(0,1)$.

Hence

$$E(S|R_5) - P[CS|R_5] = \int_{-\infty}^{\infty} \Phi[\frac{\delta}{c_5} x - (1 - \frac{\delta}{c_5}) B\sqrt{n}] d\Phi(x)$$

$$= \Phi\left[\frac{-(1 - \frac{\delta}{c_5}) B\sqrt{n}}{\sqrt{1 + (\frac{\delta}{c_5})^2}}\right] .$$

Let $E(S|R_5) - P[CS|R_5] = \varepsilon > 0$, we obtain

$$(2.42) \qquad (\frac{1}{c_5} - \frac{1}{\delta})\sqrt{n} = \sqrt{\frac{1}{\delta^2} + \frac{1}{c_5^2}} \cdot \frac{\Phi^{-1}(\varepsilon)}{B}$$

Note that

$\inf_{\Omega} P(CS|R_5] = P[T_1' \geq c_5 T_2']$, where T_1' and T_2' are defined as above.

$$P[T_1' \geq cT_2'] = P[\frac{T_1' - u'}{\sigma} \geq c_5 \left[\frac{T_2' - u''}{\sigma}\right] + (c_5 - 1)\frac{u'}{\sigma}]$$

$\approx P[Z_1 \geq c_5 Z_2 + (c_5 - 1)B\sqrt{n}]$ where Z_1, Z_2 are i.i.d. with $N(0,1)$.

Hence

$$P[T_1' \geq cT_2'] = \int_{-\infty}^{\infty} \Phi(\frac{1}{c_5} x - (1 - \frac{1}{c_5}) B\sqrt{n}) d\Phi(x) = \Phi\left[\frac{-(1 - \frac{1}{c_5}) B\sqrt{n}}{\sqrt{1 + \frac{1}{c_5^2}}}\right]$$

Setting $\inf_{\Omega} P[CS|R_5] = \varepsilon$, we obtain

$$-(1 - \frac{1}{c_5}) B\sqrt{n} = \Phi^{-1}(\varepsilon) \sqrt{1 + \frac{1}{c_5^2}}$$

$$(1 - c_5)\sqrt{n} = D\sqrt{1 + c_5^2} \quad \text{where} \quad D = \frac{\Phi^{-1}(p^*)}{B} .$$

We see that $c_5 \approx 1 - \frac{\sqrt{2}\,D}{\sqrt{n}}$ and $\frac{1}{c_5} \approx 1 + \frac{\sqrt{2}\,D}{\sqrt{n}}$. From (2.42),

$$(1 + \frac{\sqrt{2}\,D}{\sqrt{n}} - \frac{1}{\delta})\sqrt{n} = \frac{\Phi^{-1}(\varepsilon)}{B}\{\frac{1}{\delta^2} + [1 + \frac{2\sqrt{2}\,D}{\sqrt{n}} + \frac{2D^2}{n}]\}^{\frac{1}{2}}$$

$$\sqrt{n}(1 - \frac{1}{\delta}) + \sqrt{2}\,D \approx \frac{\Phi^{-1}(\varepsilon)}{B}\{\frac{1}{\delta^2} + 1\}^{\frac{1}{2}}$$

Thus

$$(2.43) \qquad n_{R_5}(\varepsilon) \approx \{\frac{\Phi^{-1}(\varepsilon)\sqrt{1+\delta^2} - \sqrt{2}\,\delta\Phi^{-1}(p^*)}{B(1-\delta)}\}^2 .$$

From (2.29) and (2.43),

$$(2.44) \qquad ARE(R_1, R_5; \delta) = \lim_{\varepsilon \to 0} \frac{n_{R_1}(\varepsilon)}{n_{R_5}(\varepsilon)} = \frac{\alpha\bar{\alpha}B^2}{a_\alpha^2 h^2(a_\alpha)}$$

If we assume that $G(x) = x$ for $0 < x < 1$, then

$$(2.45) \qquad ARE(R_1 R_5; \delta) = \frac{B^2(1-\alpha)}{\alpha} = \frac{3(1-\alpha)(2-\alpha)^2}{3(1-\alpha)(2-\alpha)^2 + \alpha^2} < 1$$

$ARE(R_1, R_5; \delta)$ is a decreasing function of α and for $\alpha = \frac{1}{2}$, it is equal to 0.931.

Note that in (2.45), R_1 is based on r-th order statistic and R_5 is based on the total life statistic with r - th failure.

(C) <u>Selection procedure for distribution $\underset{c}{<}$ ordered with respect to Weibull distribution</u>

Assume that the specified distribution $G(x)$ is given by

$$G(x) = \begin{cases} 1 - e^{-\lambda x^\alpha} & \text{for } x \geq 0 \\ 0 & \text{for } x < 0 \end{cases}$$

where $\lambda > 0$ and attention is restricted to $\alpha \geq 1$ which is assumed known. In this case, we use T_i^* as our statistic where

$$T_i^* = \sum_{j=1}^{r-1} X_{i;j,n}^{\alpha} + (n-r)X_{i;r,n}^{\alpha}, \quad i = 1,\ldots,k,$$

(as before, $X_{i;j,n}$ denote the j-th order statistic from F_i, $i = 1,\ldots,k$). Since $G(x)$ is convex with respect to the exponential distribution if $\alpha \geq 1$ and since the convex ordering is transitive, the family of distributions which are convex with respect to Weibull ($\alpha \geq 1$) will have IFR distribution. Thus our interest here is in a special subclass of IFR distributions. The rule for selecting the population which is associated with $F_{[k]}$ is as follows,

R_6: Select population π_i if and only if

$$(2.46) \qquad T_i^* \geq c_6 \max_{1 \leq j \leq k} T_j^*$$

where c_6 $(0 < c_6 < 1)$ is determined so as to satisfy the basic probability requirement.

Using the fact that if $F \underset{c}{<} G$ and $F(0) = G(0) = 0$ then $F_\alpha \underset{c}{<} G_\alpha$ for $\alpha \geq 1$, where F_α is the c.d.f. of X^α, $F(x)$ is the c.d.f. of X, G_α is the c.d.f. of Y^α and $G(y)$ is the c.d.f. of Y. Also, $G_\alpha^{-1}F_\alpha(X_{i,n}^\alpha)$ is stochastically equivalent to the i-th order statistic from $G^*(x) = 1 - e^{-x}$, for $x \geq 0$, where $X_{1,n} \leq \cdots \leq X_{n,n}$ are order statistics from F. In a manner similar to the proof of Theorem 2.1, one can prove the following theorem.

Theorem 2.5. If $F_i(x) \geq F_{[k]}(x)\ \forall x$ and $i = 1,\ldots,k$,

$F_{[k]}(0) = 0$, $F_{[k]} \underset{c}{<} G$, $G(x) = 1 - e^{-\lambda x^{\alpha}}$, $x > 0$, $\lambda > 0$ and $\alpha (\geq 1)$

is known and $n \geq \max\{r, \frac{r-1}{1-c_6}\}$, then

$$\inf_{\Omega} P[CS|R_6] = \int_0^{\infty} H^{k-1}\left(\frac{x}{c_6}\right) dH(x) \tag{2.47}$$

where $H(x)$ is the c.d.f. of a χ^2 random variable with $2r$ d.f.

(D) Selection with respect to the means for Gamma populations

Let $\pi_1,\ldots,\pi_k$ be k populations with densities

$$f_i(x) = \frac{\beta^{\alpha_i}}{\Gamma(\alpha_i)} x^{\alpha_i - 1} e^{-\beta x} \ , \quad x \geq 0 \ , \quad \beta > 0 \ , \quad \alpha_i \geq 1 \ , \quad i = 1,\ldots,k \ .$$

Let $F_i(x)$ be the distribution function of π_i , $i = 1,\ldots,k$. We are given a sample of size n from each π_i . Let T_i^* be total life statistic until r-th failure from π_i . Let $\alpha_{[1]} \leq \cdots \leq \alpha_{[k]}$ be the ordered values of α_i's . We are interested in selecting the population with the largest value $\alpha_{[k]}$ (unknown). Since the mean of π_i is $\frac{\alpha_i}{\beta}$, selection of the population with largest mean is equivalent to selecting the population with largest value, $\alpha_{[k]}$. The subset selection rule based on T_i is:

R_7: Select population π_i if and only if

$$(2.48) \qquad T_i^* \geq c_7 \max_{1 \leq j \leq k} T_j^* ,$$

where c_7 $(0 < c_7 < 1)$ is the largest value chosen to satisfy $P[CS|R_7] \geq P^*$.

Since the rule R_7 is scale invariant, we can assume $\beta = 1$.

Case 1: All α_i are unknown and ≥ 1. Let $\Omega_1 = \{\underline{\alpha} = (\alpha_1, \ldots, \alpha_k): \alpha_i \geq 1 \ \forall i\}$. In this case, by Corollary 2.2 and $F_i \underset{c}{<} G(x) = 1 - e^{-x}$, $x \geq 0$, $i = 1, \ldots, k$ we have the following result.

If $n \geq \max\{r, \frac{r-1}{1-c_7}\}$, then $\inf_{\Omega_1} P[CS|R_7] = \int_0^\infty H^{k-1}(\frac{x}{c_7})\, dH(x)$,

where $H(x)$ is the c.d.f. of a χ^2 r.v. with $2r$ d.f.

Case 2: α_i are unknown but assume $1 \leq \alpha_i \leq \Delta$, $i = 1, \ldots, k$ and Δ is known.

Let $F_\Delta(x)$ be the c.d.f. of X with density function

$$f_\Delta(x) = \frac{1}{\Gamma(\Delta)} x^{\Delta-1} e^{-x}, \quad x > 0 .$$

Let $H(x)$ be the c.d.f. of a χ^2 r.v. with $2r$ d.f. and let $h(x)$ be its density function. The following theorem is the lower bound for the probability of correct selection without any condition on n.

<u>Theorem 2.6</u>.

$$(2.49) \qquad P[CS|R_7] \geq \int_0^\infty H^{k-1}\left(\frac{2n}{c_7} x\right) \frac{2nh(2ny)}{f_\Delta(y)} e^{-x}\, dx ,$$

where $$y = F_{\Delta}^{-1}(1 - e^{-x}) \ .$$

Proof. $$P[CS|R_7] = P[T^*_{(k)} \geq c_7 \max_{1 \leq j \leq k-1} T^*_{(j)}] \ ,$$

where $T^*_{(i)}$ is associated with $\alpha_{[i]}$, $i = 1,\ldots,k$.

Since $F_{\Delta}(x) \leq F_i(x) \leq G(x) = 1 - e^{-x}$.

$$(2.50) \qquad P[CS|R_7] \geq P[T^{**}_k \geq c_7 \max_{1<j<k-1} T^{**}_j]$$

where T^{**}_k is the total life statistic until r-th failure from $G(x)$ and $T^{**}_j (j = 1,\ldots,k-1)$ is the total life statistic until r-th failure from $F_{\Delta}(x)$.

Since $\Delta \geq 1$ then $F_{\Delta} \underset{c}{<} G$. Let $\varphi(x) = G^{-1}F_{\Delta}(x)$

$$(2.51) \quad P[T^{**}_k \geq c_7 T^{**}_j \ , \ j = 1,\ldots,k-1] = P[\varphi(\tfrac{1}{n} T^{**}_k) \geq \varphi(\tfrac{c_7}{n} T^{**}_j) \ \ j = 1,\ldots,k-1]$$

By Lemma 4.1 of Barlow and Proschan [5] with $a_1 = \ldots = a_{r-1} = \frac{c_7}{n}$, $a_r = \frac{(n-r+1)c_7}{n}$, $a_i = 0$ for $i \geq r+1$ and $\varphi(X) \underset{st}{=} Y$ where $X(Y)$ is a r.v. with distribution function $F_{\Delta}(G)$ respectively, we have

$$(2.52) \quad P[\varphi(\tfrac{1}{n} T^{**}_k) > \varphi(\tfrac{c_7}{n} T^{**}_j) \ , \ j = 1,\ldots,k-1] \geq P[\varphi(\tfrac{1}{n} T^{**}_k) > \tfrac{c_7}{2n} Y_j \ , \ j = 1,\ldots,k-1]$$

where Y_j $(j = 1,\ldots,k-1)$ is a r.v. with χ^2 with $2r$ d.f.

From (2.50), (2.51) and (2.52), we have

$$P[CS|R_7] > \int_0^\infty H^{k-1}(\frac{2n}{c_7} x)\,dB(x) \text{ , where } B(x) = P[\varphi(\frac{1}{n} T_k^{**}) < x]$$

Since $\varphi(x) = -\ln(1 - F_\Delta(x))$, then $\varphi^{-1}(x) = F_\Delta^{-1}(1 - e^{-x})$. Thus

$$B(x) = P[T_k^{**} \le n\,\varphi^{-1}(x)] = H(2n\,\varphi^{-1}(x)) = H[2n\,F_\Delta^{-1}(1 - e^{-x})]$$

Now $\dfrac{dB(x)}{dx} = \dfrac{2nh(2ny)}{f_\Delta(y)} e^{-x}$ where $y = F_\Delta^{-1}(1 - e^{-x})$

Hence $$\int_0^\infty H^{k-1}(\frac{2n}{c_7} x)\,dB(x) = \int_0^\infty H^{k-1}(\frac{2n}{c_7} x) \frac{2nh(2ny)}{f_\Delta(y)} e^{-x} dx .$$

This completes the proof.

Let S denote the size of the selected subset. The expected value of S when R_7 is used is given by

$$(2.53) \qquad E(S|R_7) = \sum_{i=1}^{k} P[T_i^* \ge c_7 \max_{1<j\le k} T_j^*] .$$

Let $\Omega' = \{\underline{\alpha} = (\alpha_1,\ldots,\alpha_k):\ 1 \le \alpha_i < \Delta\ ,\ i = 1,\ldots,k\}$. For $\underline{\alpha} \in \Omega'$, since $F_\Delta(x) < F_i(x) < G(x) = 1 - e^{-x}$, then

$$E(S|R_7) \le k\ P[T_1^{**} \ge c_7 \max_{2\le j\le k} T_j^{**}]$$

where T_1^{**} is the total life statistic until r-th failure from $F_\Delta(x)$ and $T_j^{**} (j = 2,\ldots,k)$ is the total life statistic until r-th failure from $G(x)$. Hence $E(S|R_7) \le$ $k\ P[T_1^{**} \ge c_7 \max_{2\le j\le k} T_j^{**}]$. Thus

$$(2.54) \qquad \sup_{\Omega'} E(S|R_7) = k \int_0^\infty H^{k-1}\left(\frac{2x}{c_7}\right) dS(x)$$

where $H(x)$ is the c.d.f. of χ^2 r.v. with $2r$ d.f. and $S(x)$ is the c.d.f. of the total life statistic until r-th failure from $F_\Delta(X)$.

Remark 2.3. (i) We can show that the lower bound for case 2 in Theorem 2.6 is less than or equal to the lower bound for case 1.

(ii) Now we are dealing with the problem in case 2. Let $\int_0^\infty H^{k-1}\left(\frac{x}{c_7}\right) dH(x) = P^*$, then c_7 can be determined. If $n \geq \max\{r, (r-1)/(1-c_7)\}$, then we should use the lower bound for case 1. If $r \leq n < (r-1)/(1-c_7)$, then the lower bound for case 1 cannot be applied. In this case, we can use the lower bound for case 2.

(iii) Sometimes, the distribution function $S(x)$ which is defined above the remark 2.3 is hard to compute. From

$$E(S|R_7) \leq k\,P[T_1^{**} \geq c_7\, T_j^{**},\ j = 2,\ldots,k] \quad \text{where}$$

T_1^{**} is the total life statistic until r-th failure from F_Δ and T_j^{**} $(j = 2,\ldots,k)$ is the total life statistic until r-th failure from $G(x)$. Using the similar arguments in the proof of Theorem 2.6, we can get

$$E(S|R_7) \leq k \int_0^\infty H^{k-1}\left[\frac{2n}{c_7} F_\Delta^{-1}\left(1 - e^{-\frac{x}{2n}}\right)\right] dH(x)$$

where $H(x)$ is the c.d.f. of a χ^2 r.v. with $2r$ d.f. In this case, the upper bound of $E(S|R_7)$ can be computed.

3. Selecting a best population - using indifference zone approach.

Let $\pi_1, \ldots, \pi_k$ be k populations. The random variable X_i associated with π_i has an absolutely continuous distribution F_i. We assume there exists a $F_{[k]}(x)$ such that $F_{[i]}(x) \geq F_{[k]}(\frac{x}{\delta})$ for all x, $i = 1, \ldots, k-1$ and δ $(0 < \delta \leq 1)$ is specified. Let

(3.1) $\Omega(\delta) = \{\underline{F} = (F_1, \ldots, F_k) : \exists \text{ a } j \text{ such that } F_i(x) \geq F_j(\frac{x}{\delta}) \forall i \neq j\}$.

The correct selection is the choice of any population which is associated with $F_{[k]}$. We propose the selection rule

R_8 : Select population π_i if and only if

(3.2) $T_i = \max\limits_{1 \leq j \leq k} T_j$ where T_i is defined as in Section 1.2 .

We want the $P[CS|R_8] \geq P^*$, for all $\underline{F} \in \Omega$, where P^* $(\frac{1}{k} < P^* < 1)$ is specified.

<u>Theorem 3.1</u>. If $F_i \in \mathcal{F}, G \in \mathcal{G}, i = 1, \ldots, k$, $F_{[k]} \underset{c}{<} G$, $G^{-1}(0) \leq 0$, $a_j \geq 0$, $j = 1, \ldots, r$, $gG^{-1}(0) < 1$ and $a_r \geq \delta$, then

(3.3) $$P[CS|R_8] \geq \int_0^\infty G_T^{k-1}(\frac{x}{\delta})\, dG_T(x)$$

where $G_T(x)$ is the c.d.f. of T.

Proof. $P[CS|R_8] = P[T_{(k)} \geq \max_{1<j\leq k} T_{(j)}]$.

Since $F_{[i]}(\delta x) \geq F_{[k]}(x)$, $i = 1,\ldots,k-1$ and by Lemma 2.1, then

$$P[CS|R_8] = P[T_{(k)} > \delta \frac{T_{(j)}}{\delta} \ \forall\, j \neq k] > P[T_{(k)} \geq \delta\, T_j^* \ \forall\, j \neq k]$$

where $T_1^*,\ldots,T_{k-1}^*$, $T_{(k)}$ are i.i.d. with c.d.f. $W_k(x)$. Using the same argument as in Theorem 2.1, we have our theorem.

Remark 3.1. $\inf_{\Omega(\delta)} P[CS|R_8] = \int_0^\infty G_T^{k-1}(\frac{x}{\delta})\, dG_T(x)$ if $G^{-1}(0) = 0$.

For given k, δ, P^* and $G(x)$, we can possibly find the values of the pair (n,r), $(n \geq r)$ which satisfy

$$(3.4) \qquad a_r \geq \delta \quad \text{and} \quad \int_{G^{-1}(0)}^{\infty} G_T^{k-1}(\frac{x}{\delta})\, dG_T(x) \geq P^* .$$

If $G(x) = x$ for $0 < x < 1$, we can always find the values of the pair (n,r), $(n \geq r)$ which satisfy

$$n \binom{n-1}{r-1} \int_0^\infty \left[\sum_{i=r}^{n} \binom{n}{i} \left(\frac{x}{\delta}\right)^i \left(1 - \frac{x}{\delta}\right)^{n-i} \right]^{k-1} x^{r-1}(1-x)^{n-r}\, dx \geq P^*$$

If $G(x) = 1 - e^{-x}$ for $x \geq 0$, we can find the smallest integer r, say r_0, which satisfies

$\int_0^\infty H^{k-1}(\frac{x}{\delta})dH(x) \geq P^*$ where $H(x)$ is the c.d.f. of a χ^2 random variable with 2r d.f. Since $\frac{1}{\delta} a_r \geq 1$ iff $n \geq \frac{r-1}{1-\delta}$, we can find the minimum n satisfying $n \geq \max\{r, \frac{r-1}{1-\delta}\}$.

4. Selection procedure for distribution $\underset{*}{<}$ ordered with respect to G

Let $t(>0)$ be a given number. Let $N_i(t)$ be the number of failures in time t of the total n units on life test (without replacement) from π_i which has a continuous distribution F_i, $i = 1,\ldots,k$. We assume that

(4.1) $$N_i(t) \geq 1, \quad i = 1,\ldots,k,$$

and there exists a $F_{[k]}(x)$ such that $F_i(x) \geq F_{[k]}(x)$ for all x, $i = 1,\ldots,k$. The correct selection is the choice of any population which is associated with $F_{[k]}$. Let T_i be the total life statistic unitl $N_i(t)$-th failure from population π_i. Let $T_{(i)}$ be associated with $F_{[i]}$. We propose the rule

(4.2) $$R_9: \text{ Select } \pi_i \Leftrightarrow T_i \geq c_9 \max_{1 \leq j \leq k} T_j$$

where $c_9 (0 < c_9 < 1)$ is chosen so that $P[CS|R_9] \geq P^*$.

Theorem 4.1. If $F_i(x) \geq F_{[k]}(x)\ \forall x$, $i = 1,\ldots,k$ and $F_{[k]} \underset{*}{<} G$, then

(4.3) $$P[CS|R_9] \geq \int_0^\infty A_1^{k-1}\left(\frac{x}{c_9}\right) dA_2(x)$$

where $A_1(x)\,(A_2(x))$ is the c.d.f. of total life statistic until n-th (first) failure from G, respectively.

Proof. From (4.1), we have

$$P[CS|R_9] = P[T_{(k)} \geq c_9 \max T_{(j)}] \geq P[T_k^* \geq c_9 T_j^* \quad \text{for} \quad j = 1,\ldots,k-1]$$

where $\frac{T_j^*}{n}$ is the sample mean of size n from $F_{[i]}$, $i = 1,\ldots,k-1$ and $\frac{T_k^*}{n}$ is the first order statistic of size n from $F_{[k]}$.

Let

(4.4) $$T_j^{**} = \sum_{\ell=1}^{n} X_{j\ell} \quad \text{for} \quad j = 1,\ldots,k-1 .$$

where $X_{j1},\ldots,X_{jn}$ are i.i.d. from $F_{[k]}$, $j = 1,\ldots,k-1$.

since $F_{[j]}(x) > F_{[k]}(x)$ then $T_j^{**} \underset{st}{\geq} T_j^*$. Hence

(4.5) $$P[T_k^* > c_9 T_j^*,\quad j \neq k] > P[T_k^* > c_9 T_j^{**},\quad j \neq k] .$$

Let $\varphi(x) = G^{-1}F_{[k]}(x)$.

(4.6) $$P[T_k^* \geq c_9 T_j^{**},\quad j \neq k] = P[\varphi(\frac{T_k^*}{nc_9}) \geq \varphi(\frac{1}{n} T_j^{**}),\quad j \neq k]$$

Since φ is starshaped, then

(4.7) $$\varphi\left(\frac{1}{c_9} \cdot \frac{T_k^*}{n}\right) \geq \frac{1}{c_9}\,\varphi\left(\frac{T_k^*}{n}\right) .$$

By Lemma 3.1 of Barlow and Proschan [5],

$$(4.8) \qquad \varphi\left(\frac{1}{n} T_j^{**}\right) \le \sum_{\ell=1}^{n} \frac{1}{n} \varphi(X_{j\ell}) .$$

Note that $\varphi\left(\frac{T_k^*}{n}\right) \underset{st}{=} Y_k$ where Y_k is the first order statistic of size n from G and $\varphi(X_{j\ell})$ has distribution G, $j = 1,\ldots,k-1$, $\ell = 1,\ldots,n$. Let $Y_j = \sum_{\ell=1}^{n} \varphi(X_{j\ell})$, $j = 1,\ldots,k-1$. From (4.7) and (4.8), the right-hand side of (4.6) is greater than or equal to

$$\begin{aligned} P\left[\frac{1}{c_9} Y_k \ge \frac{1}{n} Y_j ,\ j = 1,\ldots,k-1\right] &= \\ &= P\left[\frac{1}{c_9} (nY_k) \ge Y_j ,\ j = 1,\ldots,k-1\right] \\ &= \int_0^\infty A_1^{k-1}\left(\frac{x}{c_9}\right) dA_2(x) . \end{aligned}$$

This completes the proof.

Acknowledgement.

The authors are very thankful Mr. Woo-Chul Kim of Purdue University for critically reading this paper and making several helpful comments and suggestions leading to an improvement of the paper.

REFERENCES

[1] Barlow, R. E. and Doksum, K. A. (1972). Isotonic tests for convex orderings. Proc. Sixth Berkeley Symp. on Math. Statist. and Prob. Vol. I, 293-323.

[2] Barlow, R. E., Esary, J. D. and Marshall, A. W. (1966). A stochastic characterization of wear-out for components and systems. Ann. Math. Statist., 37, 816-825.

[3] Barlow, R. E. and Gupta, S. S.,(1969). Selection procedures for restricted families of probability distributions. Ann. Math. Statist., 40, 905-917.

[4] Barlow, R. E., Marshall, A. W. and Proschan, F. (1963). Properties of probability distributions with monotone hazard rate. Ann. Math. Statist., 34, 375-389.

[5] Barlow, R. E. and Proschan, F. (1966). Inequalities for linear combinations of order statistics from restricted families. Ann. Math. Statist., 37, 1574-1592.

[6] Barlow, R. E. and Proschan, F. (1966). Tolerance and confidence limits for classes of distributions based on failure rate. Ann. Math. Statist., 37, 1593-1601.

[7] Barlow, R. E. and Van Zwet, W. R. (1970). Asymptotic properties of isotonic estimators for the generalized failure rate function. Part I: strong consistency. Nonparametric Techniques in Statistical Inference (ed. Puri, M. L.) 159-173. Cambridge University Press, London.

[8] Gupta, S. S. (1963). On a selection and ranking procedure for gamma populations. Ann. Inst. Statist. Math., 14, 199-216.

[9] Gupta, S. S. and McDonald, G. C. (1970). On some classes of selection procedures based on ranks. Nonparametric Techniques in Statistical Inference (ed. Puri, M.L.) 491-514, Cambridge University Press, London.

[10] Gupta, S. S. and Panchapakesan, S. (1974). Inference for restricted families: (a) multiple decision procedures (b) order statistics inequalities. Reliability and Biometry (Ed. F. Proschan and R. J. Serfling), SIAM, Philadelphia, pp. 503-596.

[11] Gupta, S. S. and Panchapakesan, S. (1975). On a quantile selection procedure and associated distribution of ratios of order statistics from a restricted family of probability distributions. Reliability and Fault Tree Analysis (Ed. R. E. Barlow, J. B. Fussell, and N. D. Singpurwalla), SIAM, Philadephia, pp. 557-576.

[12] Gupta, S. S. and Sobel, M. (1958). On the distribution of a statistic based on ordered uniform chance variables. Ann. Math. Statist., 29, 274-281.

[13] Lawrence, M. J. (1975). Inequalities of s-ordered Distributions. Ann. Statist., 3, 413-428.

[14] Lehmann, E. L. (1959). Testing Statistical Hypotheses. John Wiley, New York.

[15] Patel, J. K. (1976). Ranking and selection of IFR populations based on means. J. Amer. Statist. Assoc., 71, 143-146.

[16] Santner, T. J. (1975). A restricted subset selection approach to ranking and selection problems. Ann. Statist., 3, 334-349.

[17] Van Zwet, W. R. (1964). Convex Transformations of Random Variables. Mathematical Center, Amsterdam.

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER Mimeograph Series #518	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Subset Selection Procedures for Restricted Families of Probability Distributions		5. TYPE OF REPORT & PERIOD COVERED Technical 6. PERFORMING ORG. REPORT NUMBER Mimeo. Series #518
7. AUTHOR(s) Shanti S. Gupta and Ming-Wei Lu		8. CONTRACT OR GRANT NUMBER(s) ONR N00014-75-C-0455
9. PERFORMING ORGANIZATION NAME AND ADDRESS Purdue University Department of Statistics West Lafayette, IN 47907		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research Washington, DC		12. REPORT DATE November 1977 13. NUMBER OF PAGES 34
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release, distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Selection Procedures, Reliability, Convex, Star-shaped, IFR, IFRA, Total Life Statistic, Probability of Correct Selection.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

studies a

In this paper ~~we are interested in studying~~ multiple decision procedure for $k(k \geq 2)$ populations which are themselves unknown but which one assumed to belong to a restricted family. We propose to study a selection procedure for distributions associated with these populations which are convex-ordered with respect to a specified distribution G assuming there exists a best one. The procedure described here is based on a statistic $T_i = \sum_{j=1}^{r} a_j X_{i;j,n}$ for

DD FORM 1473, 1 JAN 73 EDITION OF 1 NOV 65 IS OBSOLETE

S/N 0102-014-6601

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

$i = 1,\ldots,k$ where $X_{i;j,n}$ is the j-th order statistic from F_i, r is a fixed positive integer ($1 \le r \le n$), $a_j = gG^{-1}(\frac{j-1}{n}) - gG^{-1}(\frac{j}{n})$ for $j=1,\ldots,r-1$, $a_r = gG^{-1}(\frac{r-1}{n})$ and g is the density of G. This statistic T_i was considered by Barlow and Doksum (1972). If $G(x) = 1 - e^{-x}$ for $x > 0$, then $nT_i = X_{i;1,n} + \ldots + X_{i;r-1,n} + (n-r+1)X_{i;r,n}$ is the total life statistic until r-th failure from F_i. This shows that the procedure based on T_i generalizes Patel's result (1976) for the IFR family.

The infimum of the probability of a correct selection is obtained and the asymptotic expression is also obtained using the subset selection approach. Some other properties of this procedure are discussed. We also study the asymptotic relative efficiencies of this rule with respect to some selection procedures proposed by Barlow and Gupta (1969) for the star-shaped ordered distributions, Gupta (1963) for the gamma populations with unknown shape parameters and etc. An example is given to illustrate the use of the selection procedure for the two independent uniform distributions. Application to quantile selection rules for distributions convex ordered with respect to Weibull distribution is given. A selection procedure for selecting the best population using the indifference zone approach is also studied.